Homework 1 comments

1. **Comment on batch size:**

Batch Size	Iterations	Learning Rate	Training Time(secs)	Accuracy
1	100000	0.0001	25.5	0.8599
10	10000	0.001	13	0.8648
100	1000	0.01	11.5	0.8599
1000	1000	0.01	14	0.862

- With the decrease in batch size, updates in SGD become less accurate to reach a minimum loss.
- To cope with this we take the smaller steps in the direction given by batch avoiding going far away from minima.
- Since the step size is smaller we need more iterations to reach the minima.
- As opposed to this the bigger batch sizes correspond to a better direction and hence larger steps can be taken in the direction given by them.
- This gives the convergence within a relatively smaller number of iterations.
- Smaller batches tend to be faster in calculating and updating the parameters in each step. On the other hand, bigger batches are slower to calculate each step.
- If the batch is too small, it will take a long time due to the large number of iterations required. If the batch size is too large it will take a long time in taking each step. The best batch lies between both of those extremities as in this case batch size 100 is the fastest to converge of all.

2. **Comment on the accuracy as a function of dataset size.**

Dataset size	Final Avg Training Loss	Test Accuracy
100	0.0033	0.6702
500	0.014	0.804
1000	0.017	0.8378
10000	0.020	0.8593

- We can see that the average training loss is well minimized for smaller datasets but the accuracy is better achieved with the bigger dataset.
- The dataset is more likely to be far from the original distribution of the data. Hence the minima calculated over the smaller dataset might not necessarily be the minima for the original distribution.

- Given the smaller loss but low accuracy, we can say that with the smaller dataset we are more likely to overfit the model.